International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-3 Issue-7, December 2013

Classification of Users in Online Video Social Networks

Kondra Mohan Raju, E. Madhukar

Abstract:- There are many online video social networks, in which Youtube is the most popular. These networks provide users to upload their own videos respect to a particular discussion. The feature provided by the networks gives the user to upload any kind of content. This creates polluted content into the system. For Example, Spammers may upload unrelated content as response to popular which increases the count of view. There is another kind of users called promoters, will gain visibility to a particular content by uploading many number of responses to increase the rank of the video. By promoting this, video will appear top in the list. This kind of activities may jeopardize the trust of the users, and social network may fail to provide genuine content. To avoid such kind of activities, we are coming up to detect the spammers and promoters. In our system we built a system same as youtube functionality having users with classification as legitimate, promoters and spammers. To distinguish between the users we allow for content and characterization attributes. These attributes can help in classifying user class. To classify the users we may use supervised classification theory. The theory is implemented on test collection. This approach successfully classified the majority of the prompters and some of the legitimate users misclassified. And most of the spammers detected form legitimate users as distinguishing is hard difficult.

Index Terms-- social network, promoters, spammers, video sharing, classification.

I. INTRODUCTION

These days video sharing networks rapidly increasing, the web has become important impart of multimedia. Youtube is the most popular online video sharing social network where massive number of videos sharing across. An example illustrates that in May 2008, 74% of the U.S social networks users are viewed 12 billion videos in that 34% of videos are form Youtube[1]. By allowing the users to publicize and upload online content, the online video social networks may become the witness of venomous and expedient user actions. Basically these systems provide three basic mechanisms for video extraction: (1) a search system. (2) List of videos with top rank. (3) the relation between user and the video. These mechanisms lead to polluting the system. The video networks can be fooled by uploading irrelevant and misleading content into the system. Some users share unrelated video as response to a topic which leads to increase in views of a video. These users we call them as spammers. The prompters and spammers generally pollute the system to generate spread to generate sales and to endanger the system popularity.

Manuscript published on 30 December 2013.
*Correspondence Author(s)
Mr. Kondra Mohan Raju, is studying M.Tech in Software Engineering from Sreenidhi Institute of Science and Technology affiliated to Jawaharlal Technological University, Hyderabad.
Mr. E. Madhukar, is working as Associate Professor in Sreenidhi Institute of Science and Technology. He is pursuing his PhD in the field of Cloud Computing.

Users may not identify the video unless they watch the video, if it is having misleading content then user get dissatisfied and also the bandwidth utility may increase. And moreover promoter videos will appear in top of the list as they may have high ranking. In this paper we classify the prompters and spammers form legitimate users in online video sharing system. We did this by extracting a huge user data set from youtube site with thousands of users. We created a labeled collection of users. From that we manually categorized the users as legitimate and the other two(spammers, promoters), who involve in users to pollute the system. Using attributes of the user, his uploads and target responses we apply supervised learning approach to identify the users. We indentified most of the promoters and distinguishing spammers, which was hard to identify from the legitimate users.

II. RELATED WORK

Content pollution has spread across all web services like E-mails, blogs[2][3]. To identify these pollutions many techniques proposed in[4][5][6][7]. These methods depends on gaining the knowledge form the textual description of the content, associated attributes related to them and apply any classification algorithm to identify the spam[8]. Some of the approaches depend on image processing methods to identify spam's in image related mails[9]. We support user based approach of identifying the pollutants as video classification technique relates to more complicated and multimedia dependent classification which is very hard to represent. The user based classification is the robust and reliable way of classifying the content as it user textual representation and content attribute. In previous studies the test collection was very less in amount and it easily distinguished between legitimate, spammers and promoters. For this we applied binary classification methods to identify spammers. Now in this paper we take a large number of test collections and apply sophisticated classification algorithms to classify the user group.

III. USER TEST COLLECTION

In order to classify the users from the online video sharing network we take large number of test collection from the system to apply the classification algorithm. The user group consists of all three classes of user. The classification is based on content attribute.

A. Crawling Social Network

We build the exact system as the social network where all the user characteristics are acquired like content attributes, video topic, responsive user, responder user and rank of the video. With all these information we build a system same as Youtube.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

B. Making a Test Collection

Generally the test collection remains to limited number. The collection may not have the effectiveness as they may not have decent number of legitimate users. To get the effective result we consider taking a large number of test collections, which contains all three kinds of users with minimum number in count. This kind of system can easily identify the effective classified list of users.

IV. UNDERSTANDING USER ATTRIBUTES

The user attributes can be identified using textual description, we consider taking the functionalities like user attributes, content attributes and network attributes. Video attributes specifies the behavior of the video uploaded by user. This indicates the quality of the video, duration, number of views, comments received, how many times it is selected as favorite and number of shares. The second attribute represents the individual behavior of the user. We expect legitimate users to spent time on selecting friends, adding videos, subscribing content representing a particular topic. We categorize the users based on attributes like number of friends' count of videos uploaded by him, number of videos viewed, average time taken to upload video.

Third kind of user attribute is, identifying the social relationship between users through interactions. On an upload or share of a video from a user will be shared with other users to respond. From this kind of relational interactions we can identify the video content.

V. IDENTIFYING SPAMMERS AND PROMOTERS

In this level we apply a supervised classification algorithm to group the users. For this we use above mentioned attributes. This algorithm learns the classification from the past labeling of content. Our aim here is potentially find the promoters and other classes of system as first effort towards helping the system admin to identify content polluters.

In identification strategy we use evaluation metrics to classify the users. The classification strategy consists of information retrieval system metrics like precision and recall. The classification strategies can follow in flat or hierarchical manner. For classification we use Support Vector Machine(SVM) which gives the best results in classification test data. We use two kinds of classification approaches. They are Flat classification and hierarchical classification. The flat way of hierarchy represents the user set in simple manner which divides them as Promoters(P), Spammers(S) and Legitimate(L) users.

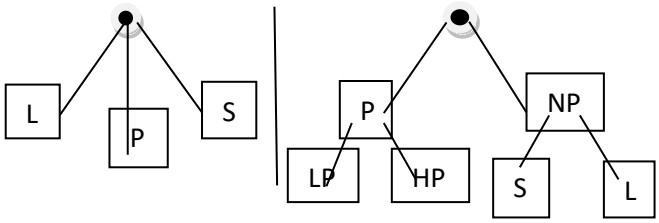

Fig 1: User Classification: Flat(Left) and Hierarchical (Right)

Considering the hierarchical way, we divide the users in fuzzy logic manner. The users may be divided as heavy promoters (HP), light promoters(LP), non-Promoters(NP).

VI. CONCLUSION

Promoters and spammers not only just pollute the content in the system but also influence the system performance. This paper gives the way to identify spammers and promoters which will help the system administrator to filter different user group. The classification techniques used here identified spammers and prompters most effectively and partially failed in the case of legitimate user set. This way of classification gives the flexibility to system administrator and he may send warning messages to user who is polluting the content.

We envision for reducing the cost of labeling process by enhancing the supervised classification theory to identify user groups. The refinement process explores to other techniques related to classification.

REFERENCES

1. comscore: Americans viewed 12 billion videos online in may2008. http://www.comscore.com/press/release.asp?press=2324.
2. L. Gomes, J. Almeida, V. Almeida, and W. Meira. Workload models of spam and legitimate e-mails. *Performance Evaluation*, 64, 2007.
3. A. Thomason. Blog spam: A review. In *Conference on Email and Anti-Spam (CEAS)*, 2007.
4. C. Castillo, D. Donato, A. Gionis, V. Murdock, and F. Silvestri. Know your neighbors: Web spam detection using the web topology. In *Int'l ACM SIGIR*, 2007.
5. Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen. Combating web spam with trustrank. In *Int'l. Conference on Very Large Data Bases (VLDB)*, 2004.
6. Y. Lin, H. Sundaram, Y. Chi, J. Tatemura, and B. Tseng. Detecting splogs via temporal dynamics using self-similarity analysis. *ACM Transactions on the Web (TWeb)*, 2, 2008.
7. Y. Xie, F. Yu, K. Achan, R. Panigrahy, G. Hulten, and I. Osipkov. Spamming botnets: Signatures and characteristics. In *ACM SIGCOMM*, 2008.
8. P. Heymann, G. Koutrika, and H. Garcia-Molina. Fighting spam on social web sites: A survey of approaches and future challenges. *IEEE Internet Computing*, 11, 2007.
9. C. Wu, K. Cheng, Q. Zhu, and Y. Wu. Using visual features for anti-spam filtering. In *IEEE Int'l Conference on Image Processing (ICIP)*, 2005.
10. Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of topological characteristics of huge online social networking services. In *Int'l World Wide Web Conference (WWW)*, 2007.

AUTHOR PROFILE

Mr. Kondra Mohan Raju, is studying M.Tech in Software Engineering from Sreenidhi Institute of Science and Technology affiliated to Jawaharlal Technological University, Hyderabad. His research area is in Cloud Computing, Image Processing, Data Mining.

Mr. E. Madhukar, is working as Associate Professor in Sreenidhi Institute of Science and Technology. He is pursuing his PhD in the field of Cloud Computing. His area of teaching related to Image Processing, Cloud Computing.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)